Automatic Voice Title Generation by Extractive Compression of Long Product Titles

A Deep Learning based approach to product title compression

Snehasish Mukherjee (mukherji@stanford.edu)
Phaniram Sayapaneni (Phaniram.Sayapaneni@walmartlabs.com)

Objective

Generating friendly, short and crisp product titles is of utmost value in voice commerce, since it directly effects the end user experience. Long product titles are not suitable for reading out during an automatic voicebased shopping related

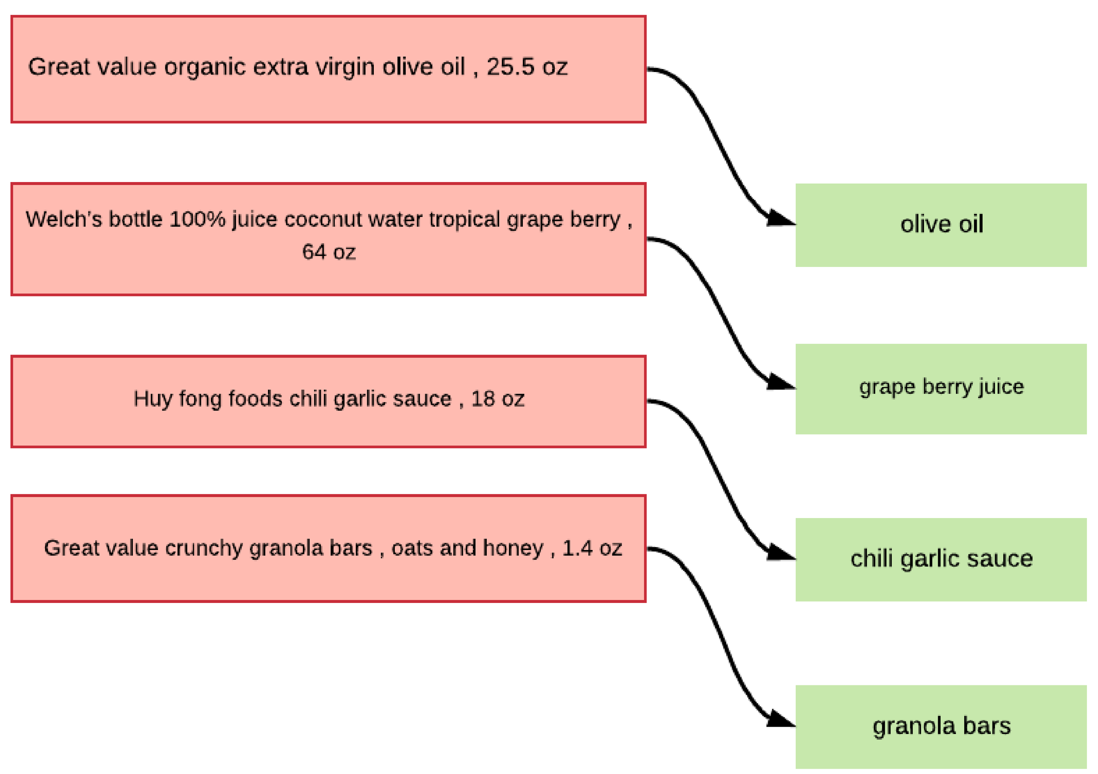


Figure 1: The long product titles in red, could be summarized with the ones in green

We solve this problem by learning the right strategies to compress long titles such that the obtained short title is still able to describe the product at it's best.

Deep Learning Approach

We model the problem of generating voice friendly titles as a supervised sequence labelling problem. The words tagged by the model as important will be part of the summary. Formally, given the input training pair (X, Y) the model tries to learn the optimal parameter θ' such that

$$\theta' = \text{argmax} \sum \log p(Y|X;\theta) \quad (1)$$

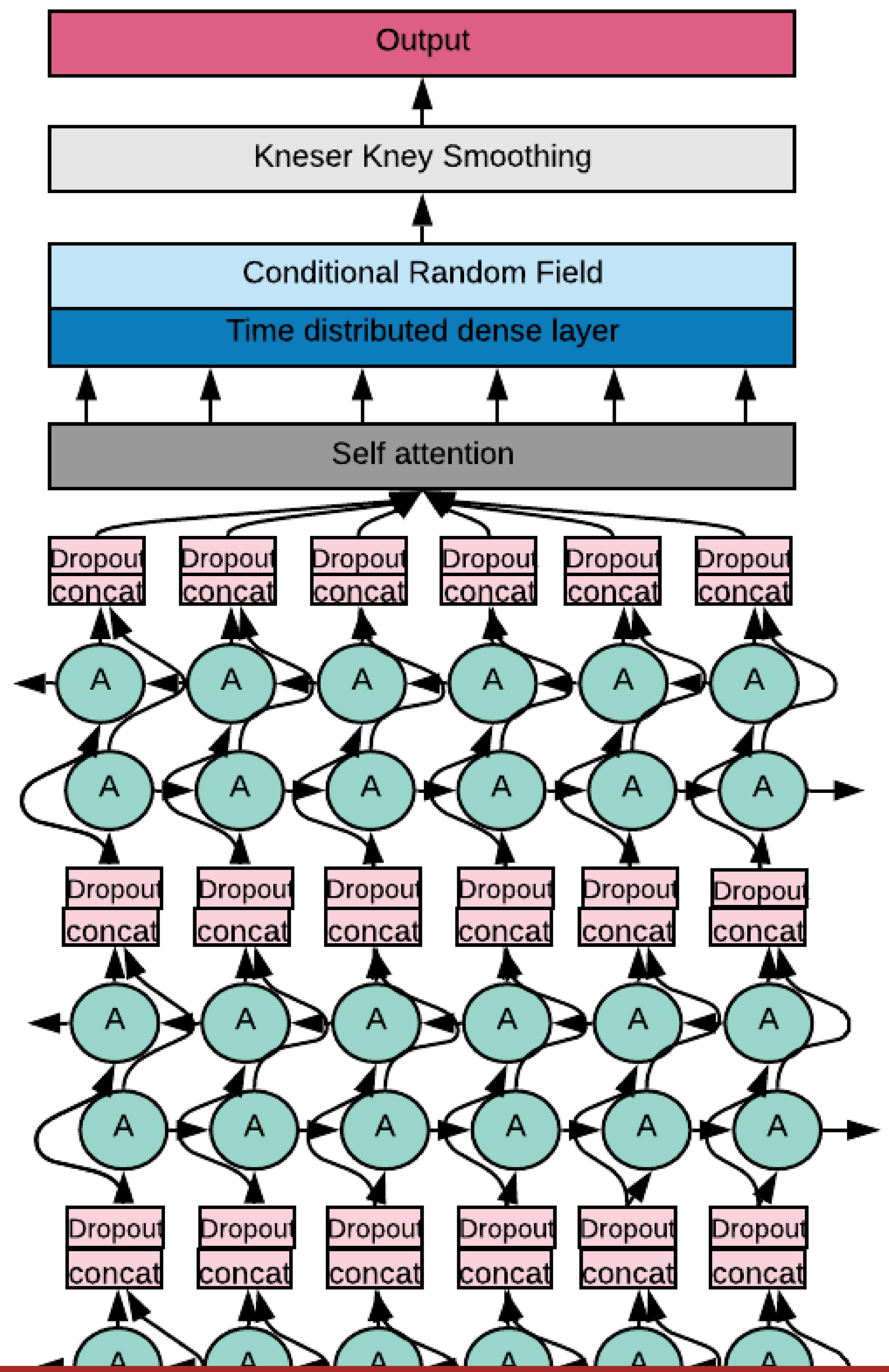
where $Y_i \in \{0, 1\}$ and $X_i \in \mathbb{R}^d$ is an embedding of a word in a d-dimensional vector space. Additionally in our problem to we keep $|X| = |Y|$ and use padding to guarantee this.

Deep Learning Approach

- We have implemented the model using a stacked Bi-LSTM based architecture.
- We have also approached the problem though transfer learning techniques, where we have taken the pre-trained state of art models and fine tuned them to our use case.

  The Bi-LSTM based approach involves learning the embeddings of words from scratch.

  While the Bi-LSTM network learns to represent the sentence, the self attention has been added to compute the context vector guiding the model to learn the dependencies across different words in a given sentence. The Conditional random field layer has been added to model the conditional distribution of the extracted short title given the word representations.
- On the other hand we have fine tuned the pretrained models to perform the downstream task, which in our case is sequence labelling. Individual model's performance as a result has been evaluated in the below section.



Results

ROUGE-1 METRICS

Model	Precision	Recall	F1-Score
3-Layer stacked LSTM	0.831	0.849	0.816
3-Layer Stacked BiLSTM	0.831	0.849	0.816
3-Layer BiLSTM with Attention	0.878	0.882	0.862
3-layer BiLSTM with CRF	0.869	0.872	0.852
BERT	0.797	0.780	0.750
OpenAI GPT-1	0.730	0.656	0.656
XLNET	0.690	0.655	0.640

Table 1: Performance across different approaches

Generated responses

Here are some of our results where model out performed human labelling, showing how model is resistant to typical human prone errors.

- **Long title:** luna whole nutrition bar , chocolate peppermint , 8g protein , 6 ct
  Human: protein
  Model: nutrition bar
- **Long title:** warheads sour candy assortment easter egg , 5.22 oz
  Human: easter egg
  Model: sour candy
- **Long title:** butterball everyday turkey sausage polska kielbasa , 13.0 oz
  Human: kielbasa
  Model: turkey sausage kielbasa

Discussions

- The results are encouraging, a simple Bi-LSTM memory based network has been able to perform well on the short title generation
- Amongst the transfer learning approaches, Bert seems to have out performed other models for our use case. This difference could be attributed to the differences in the model architecture and word tokenization techniques.
- While the supervised learning approach seems to have performed better than transfer learning approaches, it is highly limited on the vocabulary, data it has been trained on. Whereas transfer learning approach was performing much better across broad variety of products(products it has never seen in the training data).

Future Work

- We would like to explore more on the generative approaches such as GANs, VAEs moving forward. Such that we could generate short title during inference time.
- Moving forward we will try to scale up this solution to tens-of millions of product titles. We will also explore other word embeddings likeELMO and fast-text in improving the model performance.
- Finally we would like to apply reinforcement learning techniques to generate short titles in manner that our customers most like.

References

[1] Ramesh Nallapati, Bowen Zhou, Cícero Nogueira dos Santos, Çaglar Gülçehre, and Bing Xiang. Abstractive text summarization using sequence-to-sequence rnns and beyond. In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, CoNLL 2016, Berlin, Germany, August 11-12, 2016*, pages 280–290, 2016.

[2] Yang Liu. Fine-tune BERT for extractive summarization. *CoRR*, abs/1903.10318, 2019.